```
popquizbox[1] colback=nipun-lightblue!10, colframe=nipun-blue,
boxrule=2pt, arc=3pt, left=8pt, right=8pt, top=8pt,
bottom=8pt, title= Quick Quiz 1, fonttitle=,
coltitle=nipun-white, colbacktitle=nipun-blue,
enhanced, attach boxed title to top left=xshift=0pt,
yshift=-2pt, boxed title style=arc=3pt, boxrule=0pt
definitionbox[1] colback=nipun-green!8, colframe=nipun-green,
```

[BoldFont=Fira Sans SemiBold]Fira Sans Book Fira Mono

boxrule=1.5pt, arc=2pt, left=6pt, right=6pt, top=6pt, bottom=6pt, title= **Definition:** 1, fonttitle=, coltitle=nipun-white, colbacktitle=nipun-green

examplebox[1] colback=nipun-orange!8, colframe=nipun-orange,

boxrule=1.5pt, arc=2pt, left=6pt, right=6pt, top=6pt, bottom=6pt, title= **Example:** 1, fonttitle=,

coltitle=nipun-white, colbacktitle=nipun-orange

keypointsbox colback=nipun-blue!8, colframe=nipun-blue,

boxrule=1.5pt, arc=2pt, left=6pt, right=6pt, top=6pt,

bottom=6pt, title= **Key Points**, fonttitle=,

Lagrangian and Duality

Nipun Batra

June 28, 2020

IIT Gandhinagar

Lectures heavily inspired by the Maths for Machine learning book

• Minimax inequality states: $\max_{\mathbf{y}} \min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leqslant \min_{\mathbf{x}} \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$

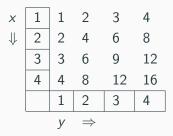
- Minimax inequality
 states:max_y min_x q(x, y) ≤ min_x max_y q(x, y)
- We first prove For all $x, y = \min_{x} q(x, y) \leqslant \max_{y} q(x, y)$

• Let us choose q(x, y) = xy

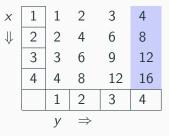
- Let us choose q(x, y) = xy
- ullet Let us first find $\max_{oldsymbol{y}} q(oldsymbol{x}, oldsymbol{y})$

- Let us choose q(x, y) = xy
- ullet Let us first find $\max_{oldsymbol{y}} q(oldsymbol{x}, oldsymbol{y})$

- Let us choose q(x, y) = xy
- Let us first find $\max_{y} q(x, y)$

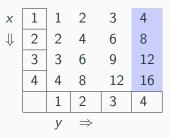


• For each value of x, we find y that maximizes q(x,y)



3

- For each value of x, we find y that maximizes q(x, y)
- y = 4 maximizes $q(x, y) \forall x$



• For each value of y, we find x that minimizes q(x, y)

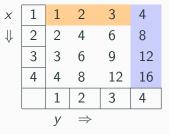
X	1	1	2	3	4
\Downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

4

- For each value of y, we find x that minimizes q(x, y)
- x = 1 minimizes $q(x, y) \forall y$

X	1	1	2	3	4
\Downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

• We just showed For all $x, y = \min_{x} q(x, y) \leqslant \max_{y} q(x, y)$

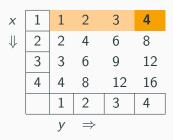


5

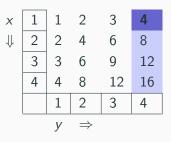
- We just showed For all $x, y = \min_{x} q(x, y) \leqslant \max_{y} q(x, y)$
- The equality occurs at x = 1, y = 4



• Let us now find $\max_{y} \min_{x} q(x, y)$

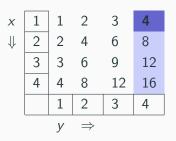


• Similarly, let us now find $\min_{x} \max_{y} q(x, y)$



7

- Similarly, let us now find $\min_{x} \max_{y} q(x, y)$
- We can thus see our Minimax inequality $\max_{\mathbf{y}} \min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leqslant \min_{\mathbf{x}} \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$



Our problem is of the form

$$\min_{m{x}} f(m{x})$$
 subject to $g_i(m{x}) \leqslant 0$ for all $i=1,\ldots,m$

Our problem is of the form

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1} (g_i(\mathbf{x}))$$

Our problem is of the form

$$\min_{m{x}} f(m{x})$$
 subject to $g_i(m{x}) \leqslant 0$ for all $i=1,\ldots,m$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\mathbf{x}))$$

where 1(z) is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

Our problem is of the form

$$\min_{m{x}} f(m{x})$$
 subject to $g_i(m{x}) \leqslant 0$ for all $i=1,\ldots,m$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\mathbf{x}))$$

where 1(z) is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

This would give infinte penalty if constraint is not satisfied. But, this formulation is hard to solve too.

Idea: Introduce Lagrange multipliers ($\lambda_i \geq 0$) to "approximate" J(x)

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

Idea: Introduce Lagrange multipliers $(\lambda_i \ge 0)$ to "approximate" J(x)

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

What is the relationship between $\mathfrak{L}(\mathbf{x}, \lambda)$ and $J(\mathbf{x})$ given $\lambda_i \geq 0$?

Idea: Introduce Lagrange multipliers $(\lambda_i \ge 0)$ to "approximate" J(x)

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

What is the relationship between $\mathfrak{L}(\mathbf{x}, \lambda)$ and $J(\mathbf{x})$ given $\lambda_i \geq 0$?

When $\lambda \geqslant 0$, the Lagrangian $\mathcal{L}(x,\lambda)$ is a lower bound of J(x). Hence, the maximum of $\mathfrak{L}(x,\lambda)$ with respect to λ is

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

9

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{\pmb{x} \in \mathbb{R}^d} \max_{\pmb{\lambda} \geqslant \pmb{0}} \mathfrak{L}(\pmb{x}, \pmb{\lambda})$$

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\mathbf{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda}) \geqslant \max_{\mathbf{\lambda} \geqslant \mathbf{0}} \min_{\mathbf{x} \in \mathbb{R}^d} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda}) \geqslant \max_{oldsymbol{\lambda} \geqslant \mathbf{0}} \min_{oldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

We can write the dual objective as a function of λ as

$$\mathfrak{D}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

• Primal objective:

$$\min_{m{x}} f(m{x})$$
 subject to $g_i(m{x}) \leqslant 0$ for all $i=1,\ldots,m$

• Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

• Or, primal objective $=J(x)\geq \max_{\lambda}\mathfrak{D}(\lambda)$

Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leq 0$ for all $i = 1, ..., m$

- Or, primal objective = $J(x) \ge \max_{\lambda} \mathfrak{D}(\lambda)$
- Or, primal objective (in terms of x) ≥ dual objective (in terms of λ)

Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

- Or, primal objective = $J(x) \ge \max_{\lambda} \mathfrak{D}(\lambda)$
- Or, primal objective (in terms of x) ≥ dual objective (in terms of λ)
- For SVM like formulations, primal objective is the same as dual objective (strong duality)

• Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leq 0$ for all $i = 1, ..., m$

- Or, primal objective = $J(x) \ge \max_{\lambda} \mathfrak{D}(\lambda)$
- Or, primal objective (in terms of x) ≥ dual objective (in terms of λ)
- For SVM like formulations, primal objective is the same as dual objective (strong duality)
- For some problems, there is a "daulity-gap" between the two objectives